Rounding of probability forecasts : The SPF forecast probabilities of negative output growth

Michael P. Clements

No 869

WARWICK ECONOMIC RESEARCH PAPERS

DEPARTMENT OF ECONOMICS

THE UNIVERSITY OF WARWICK

Rounding of probability forecasts:
The SPF forecast probabilities of negative output growth

Michael P. Clements*

Department of Economics,

University of Warwick,

UK. CV4 7AL.

m.p.clements@warwick.ac.uk.

September 22, 2008

Abstract

We consider the possibility that respondents to the Survey of Professional Forecasters round their probability forecasts of the event that real output will decline in the future. We make various assumptions about how forecasters round their forecasts, including that individuals have constant patterns of responses across forecasts. Our primary interests are the impact of rounding on assessments of the internal consistency of the probability forecasts of a decline in real output and the histograms for annual real output growth, and on the relationship between the probability forecasts and the point forecasts of quarterly output growth.

Journal of Economic Literature classification: C53, E32, E37

Keywords: Rounding, probability forecasts, probability distributions.

*Computations were performed using code written in the Gauss Programming Language.

1 Introduction

Reported forecast probabilities may sometimes be rounded in the sense that a 33% probability might be reported as 35%, or 30%, or 40%, or 50%, or even 0%, reflecting rounding to the nearest 5%, 10%, 20%, 50%, or to either 0 or 100%. In this paper we consider probability forecasts taken from one of the best known and longest-running surveys of the US economy, the Survey of Professional Forecasters (SPF), and assess the potential impact of rounding in terms of: *i*) the effect on internal consistency of the probability forecasts of a decline in real output and the histograms for annual real output growth, and *ii*) the relationship between the probability forecasts and the point forecasts of quarterly output growth. The SPF is perhaps a unique resource from the perspective of allowing us to assess the impact of rounding on the internal consistency of the survey respondents, as the SPF allows us calculate matched point, probability and probability distribution forecasts for the individual forecasters. In the case of the SPF probability forecasts we have no idea to what extent the survey participants indulge in rounding, and presumably this is true of most surveys where respondents are required to report probabilities. We do not know whether there are differences between individuals in this regard, or even whether individuals behave in the same way over time. Hence we will investigate the potential impact of rounding under a number of plausible assumptions, including the assumption that individuals adopt common response patterns, as motivated by Manski and Molinari (2008).

The SPF probability forecasts are particularly interesting from the perspective of analying reporting behaviour. Because the event of negative real output growth is relatively rare, and is perceived as such by the SPF respondents, the forecast probabilities are often small, and are sometimes reported as zero. This is especially problematic, as a reported forecast probability of zero could mask an actual assessment close to a fifty-fifty chance for 'coarse rounding', i.e., if actual assessments in the interval $[0, 50]$ are reported as 0, and those $[50, 100]$ are reported as 100%. Alternatively a reported probability of 0 could reflect an actual belief the event is certain not to occur. Moreover, a number of authors have investigated the consistency of the different types of forecasts made by the respondents, and have noted some apparent inconsistencies (Engelberg, Manski and Williams (2007), Clements (2008c, 2008b)). We investigate the extent to which some of these may be attributable to rounding of the probability forecasts.

The next section briefly describes the SPF data. We describe the different assumptions we will make about the degree of rounding behind the reported numbers, and how intervals can be generated to replace the point probability forecasts. The consequences of the different rounding assumptions in terms of the distributions of the widths of the implied intervals are presented. Section 3 reviews the approach of Clements (2008c) to determining whether the reported probability forecasts of a decline in output are consistent with their probability distributions, and calculates the impact of allowing for rounding. Section 4 investigates the impact of rounding on the relationship between the probability of decline forecasts and the output growth forecasts. Section 5 offers some concluding remarks.

2 The Survey of Professional Forecasters (SPF) data and rounding

The SPF quarterly survey began as the NBER-ASA survey in 1968:4 and runs to the present day. The survey questions elicit information from the respondents on their point forecasts for a number of variables; their histograms for output growth and inflation; and the probabilities they attach to declines in real output. The SPF point forecasts and histograms have been widely analysed in separate exercises.[1] The probability forecasts (of the event that output will decline) have received relatively little attention.[2]

We will also make use of the quarterly Real Time Data Sets for Macroeconomists (RTDSM) maintained by the Federal Reserve Bank of Philadelphia (see Croushore and Stark (2001)). This consists of a data set for each quarter that contains only those data that would have been available at a given reference date: subsequent revisions, base-year and other definitional changes that occurred after the reference date are omitted.[3] The RTDSMs tie in with the SPF surveys such that a respondent to, say, the 1995:Q1 survey would have access to the data in the Feb 1995 RTDSM. The datasets contain quarterly observations over the period 1947:Q1 to the quarter before the reference

[1] For example, the point forecasts have been analysed by Zarnowitz (1985), Keane and Runkle (1990), Davies and Lahiri (1999), and the probability distributions by Diebold, Tay and Wallis (1999), Giordani and Söderlind (2003) and Clements (2006). A detailed academic bibliography of papers that use SPF data is maintained at http://www.phil.frb.org/econ/spf/spfbib.html.

[2] Although the SPF produces an 'anxious index' by averaging the individual respondents' probabilities of declines in real output in the following quarter, and this is shown to be correlated with the NBER business cycle periods of expansion and recession. Recently Clements (2008a) examines the degree of disagreement in these forecasts across individuals.

[3] Later vintages of data would not serve our purpose, as they typically contain revisions and definitional changes that were largely unpredictable (see Faust, Rogers and Wright (2005)) based on information available at the time, and as such would not have been known to the SPF respondents.

date (1994:4, in our example), and allow us to construct measures based solely on the data vintages available to the SPF respondents at the time they made the forecast.

Our main focus is on the forecasts of a decline in real output (negative output growth), and the extent to which the figures reported by the survey respondents may be rounded. We will also make use of the histograms for annual real output growth, and the point forecasts of quarterly real output. We use data from 1981:3 to 2005:1, as prior to 1981:3 the histograms for output growth referred to nominal output, and point forecasts for real GDP (GNP) were not recorded.[4]

The probability forecasts consist of reported probabilities of a decline in output in the current quarter (the survey date quarter) relative to the previous quarter, for the next quarter relative to the current quarter, and so on up to the same quarter a year ahead relative to three quarters ahead.[5] The probability distributions refer to the annual change from the previous year to the year of the survey, as well as of the survey year to the following year, although we use only the former. The point forecasts are of the level of output in the current year, as well as for the previous quarter, for the current quarter, the next four quarters, and for the current year.[6]

The total number of usable point and probability forecasts across all surveys and respondents is 2462. These forecasts come from the 95 quarterly surveys from 1981:3 to 2005:1, and from 181 different respondents. We restrict the sample to only include regular forecasters - those who have responded to more than 12 surveys.[7] This gives 73 respondents. These regular respondents account for 1969 forecasts, some 80% of the total.

Table 1 reports the proportion of reported probabilities which are multiples of 5, i.e., reported as 0,5,10 etc. A striking feature of the table is that over a half of all forecasts are reported as multiples of 10, and that around 90% are reported as multiples of either 5 or 10 for each of the five forecast horizons. It is clear that these values are being reported far more often than chance would dictate. Also evident is that a decline in output is perceived to be a low probability event. 30%

[4] The definition of real output changes from GNP to GDP in 1992:1, and there are also base year changes over the period, but these are not expected to affect our findings.

[5] Hence for, say, the 1995 Q1 survey, respondents are asked to report the probabilities they attach to declines in 95:Q1 relative to 94:Q4, 95:Q2 relative to 95:Q1, 95:Q3 relative to 95:Q2, 95:Q4 relative to 95:Q3 and 96:Q1 relative to 95:Q4.

[6] The most recent figures are sent out with the survey forms, so that a respondent to the 95:Q1 survey would receive the values for the variables for which point forecasts are required up to and including 94:Q4. Most respondents use the 94:Q4 figures as their forecasts of that quarter.

[7] One might suppose that regular respondents are more au fait with the survey and what is required of them, so that errors in filling in the survey questionnaires are less likely.

assign a zero probability for their current quarter forecast - this proportion approximately halves for forecasts of the next quarter, and then halves again, as the proportion reporting a probability of 10% approximately doubles as the horizon increases.

Clearly, not all the forecasts reported as multiples of 5 correspond to underlying subjective probability forecasts which are exact multiples of 5. Manski and Molinari (2008) argue that whilst the common practice of ignoring the possibility that responses may be rounded leads to precise point data it is not credible. Suppose instead that probability forecasts reported as multiples of 5 i.e., 0, 5, 10, ..., 100 reflect actual probability assessments in the intervals [0,2.5], [2.5,7.5], ..., [97.5,100]. We refer to this relatively precise degree of rounding as 'M5'. Of course, responses may be round to the nearest 10. In which case all we know is that reported multiples of 10 (i.e., 0,10, ..., 100 correspond to underlying probability assessments that fall in the intervals [0,5], [5,15], ..., [95,100]. If we couple this assumption with the assumption that forecasts reported as multiples of 5 (but excluding those reported as multiples of 10) are treated as in M5, then we have strategy M10. The assumption that a grosser form of rounding occurs would be that reported multiples of 50 (i.e., 0,50,100) would actually reflect intervals of [0,25], [25,75] and [75,100], and this assumption with M10 gives M50. The grossest from of rounding behaviour would indicate that reported forecasts of 0 and 100 are associated with the intervals [0,50] and [50,100]. This is referred to as M100.

In table 2 we report the distribution of the widths of the interval forecast generated by assuming rounding of the form M5, M10, M50 and M100 for each of the five forecast horizons. Note that zero widths result when the reported forecasts are not a multiple of 5, where formally we assume that the upper and lower values of the interval are equal to the reported value. Under M5, the maximum width is 5 percentage points, corresponding to a reported point forecast which is a multiple of 5 other than 0 or 100 - boundary forecasts of 0 or 100 give rise to intervals of 2.5 under M5. If we allow for the grosser forms of rounding then the intervals naturally become less informative. As an example, under M100, just over one third (34.5%) of the current quarter forecasts have a width of 50 percentage points, predominantly reflecting reported assessments of zero being classed as intervals [0,50]. Even allowing for more modest degrees of rounding, such as rounding to the nearest 10, M10, nearly half the forecasts are replaced with intervals widths of 10 at the longer horizons.

Table 1: Reported probabilities of decline of SPF respondents

Forecast	Current quarter	1-quarter	2-quarter	3-quarter	4-quarter
0	.300	.134	.076	.071	.074
5	.132	.157	.145	.117	.119
10	.111	.174	.204	.202	.191
15	.034	.068	.115	.114	.098
20	.070	.103	.130	.153	.155
25	.031	.060	.062	.082	.082
30	.044	.060	.069	.068	.077
35	.017	.016	.022	.028	.023
40	.034	.039	.034	.028	.037
45	.010	.008	.003	.006	.007
50	.029	.034	.023	.020	.031
55	.000	.000	.000	.000	.000
60	.009	.007	.009	.006	.006
65	.003	.004	.000	.000	.001
70	.006	.004	.002	.003	.002
75	.006	.007	.005	.003	.003
80	.008	.005	.003	.001	.002
85	.001	.001	.000	.000	.000
90	.011	.002	.001	.001	.001
95	.004	.001	.000	.001	.000
100	.016	.002	.000	.000	.001
Proportion reported as:					
10 or 5	0.872	0.884	0.901	0.901	0.909
10	0.636	0.563	0.550	0.552	0.577
5 (excl. multiples 10)	0.236	0.321	0.351	0.349	0.332
No. Forecasts	1969	1969	1969	1969	1969

.

Table 2: Widths of intervals assuming the reported probabilities reflect different degrees of rounding

'Rounding'	Percentage of intervals with a width of percentage points:					
	0	2.5	5	10	25	50
Horizon: current						
M5	12.8	31.6	55.6	0.0	0.0	0.0
M10	12.8	0.0	55.2	32.0	0.0	0.0
M50	12.8	0.0	23.6	29.2	31.6	2.9
M100	12.8	0.0	23.6	29.2	0.0	34.5
'Individual'	76.7	6.4	16.5	0.3	0.0	0.0
Horizon: 1-quarter ahead						
M5	11.6	13.6	74.9	0.0	0.0	0.0
M10	11.6	0.0	45.7	42.8	0.0	0.0
M50	11.6	0.0	32.1	39.4	13.6	3.4
M100	11.6	0.0	32.1	39.4	0.0	16.9
'Individual'	57.7	5.8	35.9	0.6	0.0	0.0
Horizon: 2-quarter ahead						
M5	9.9	7.6	82.5	0.0	0.0	0.0
M10	9.9	0.0	42.8	47.3	0.0	0.0
M50	9.9	0.0	35.1	45.0	7.6	2.3
M100	9.9	0.0	35.1	45.0	0.0	10.0
'Individual'	52.7	1.7	45.1	0.4	0.0	0.0
Horizon: 3-quarter ahead						
M5	9.9	7.1	83.1	0.0	0.0	0.0
M10	9.9	0.0	42.0	48.1	0.0	0.0
M50	9.9	0.0	34.9	46.1	7.1	2.0
M100	9.9	0.0	34.9	46.1	0.0	9.1
'Individual'	52.1	2.8	45.1	0.0	0.0	0.0
Horizon: 4-quarter ahead						
M5	9.1	7.5	83.4	0.0	0.0	0.0
M10	9.1	0.0	40.7	50.2	0.0	0.0
M50	9.1	0.0	33.2	47.0	7.5	3.1
M100	9.1	0.0	33.2	47.0	0.0	10.7
'Individual'	54.6	2.7	42.7	0.0	0.0	0.0

There are 1969 forecasts. M5 to M100 indicate increasingly grosser degrees of rounding, as described in the text. 'Individual' indicates that the degree of rounding is inferred from individual-level behaviour.

In the next section we investigate the potential impact of rounding on the assessment of the consistency of different types of forecasts made by the SPF respondents, namely the consistency of the probability forecasts and the histogram forecasts. Before doing so, we use response patterns by the individual respondents to the SPF to infer the degree of rounding undertaken by each individual, as suggested by Manski and Molinari (2008). As we will show, the use of the information provided by individual response patterns is valuable in markedly narrowing the intervals that replace the reported probability forecasts.

Assume that each individual applies the same rounding rule when they respond to the SPF question on the probability of decline, but that the rounding rule may differ by individual. Adapting the algorithm proposed by Manski and Molinari (2008), we proceed as follows. If all the responses by individual j are either 0 or 100, we assume that individual j engages in gross rounding, so that responses of 0 are replaced by the interval [0,50], and responses of 100 by in the interval [50,100]. If all responses are 0, 50, 100, we assume that the respondent is rounding to the nearest 50, and these point forecasts are replaced by the following intervals, [0,25], [25,75], [75,100]. Similarly if all responses are multiples of 10, and if all responses are multiples of 5. This can be codified formally as follows, where the reported point forecasts made by j to survey t_j, p_{j,t_j}, is replaced by the interval $\left[p_{j,t_j,L}, p_{j,t_j,U}\right]$ according to the following scheme:

If $p_{j,t_j} = (0, 100)$, for all t_j, then $\left[p_{j,t_j,L}, p_{j,t_j,U}\right] = [\max\left(0, p_{j,t_j} - 50\right), \min\left(p_{j,t_j} + 50, 100\right)]$.
If $p_{j,t_j} = (0, 50, 100)$, for all t_j, then $\left[p_{j,t_j,L}, p_{j,t_j,U}\right] = [\max\left(0, p_{j,t_j} - 25\right), \min\left(p_{j,t_j} + 25, 100\right)]$.
If $p_{j,t_j} = (0, 10, 20, \ldots, 100)$, for all t_j, then $\left[p_{j,t_j,L}, p_{j,t_j,U}\right] = [\max\left(0, p_{j,t_j} - 5\right), \min\left(p_{j,t_j} + 5, 100\right)]$.
If $p_{j,t_j} = (0, 5, 10, 15, , \ldots, 100)$, for all t_j, then $\left[p_{j,t_j,L}, p_{j,t_j,U}\right] = [\max\left(0, p_{j,t_j} - 2.5\right), \min\left(p_{j,t_j} + 2.5, 100\right)]$.
Otherwise $\left[p_{j,t_j,L}, p_{j,t_j,U}\right] = \left[p_{j,t_j}, p_{j,t_j}\right]$.

Note that t_j will typically be an element of a subset of the quarterly surveys from 1981Q3 to 2005Q1. Due to the time dimension, non-responses are not assumed to provide any information on the personal response pattern, unlike in the study of Manski and Molinari (2008).

The distribution of the widths of the resulting intervals are recorded in table 2 in the rows 'Individual'. The narrowing of the intervals when individual response pattern information is used is marked. For example, for the current quarter forecasts over three quarters (76.7%) of the intervals are in fact points, and virtually no intervals are of greater width than 5 at any of the horizons.

Use of this schema effectively rules out rounding to anything other than a multiple of 5. Although there are many instances of forecasts reported as multiples of 10, 50 and 100, as evident from table 1, virtually all respondents produce at least one forecast which is a multiple of five or has a finer value, ruling out the grosser forms of rounding under the maintained assumption that individuals operate with the same rounding rule for all their forecasts.

Comparing the width distributions for the intervals calculated using individual response patterns for the current quarter and longer horizon forecasts it is apparent that the degree of rounding increases with the forecast horizon. The fraction of zero-width intervals falls from around three-quarters to close to a half. We regard this as evidence that the SPF respondents are rounding to convey 'ambiguity' rather than to 'simplify communication' (in the parlance of Manski and Molinari (2008)), because if the latter were the case there would be no reason to expect the degree of rounding to increase with the forecast horizon.

3 The consistency of the probability forecasts and the probability distributions

Clements (2008c) takes the reported probability forecasts at face value, and assesses the consistency of these forecasts and the respondents' probability distributions by adapting the bounds approach of Engelberg *et al.* (2007). Engelberg *et al.* (2007) use a bounds approach to assess the consistency of the point forecasts and probability distributions. Because the probability distribution is reported as a histogram, the survey return provides an incomplete picture, and it is only possible to calculate point measures of, say, the mean (in Engelberg et al's case) if we are prepared to make assumptions about the distribution of the probability mass within the histogram intervals. If we do not wish to make such assumptions we can still bound the values that the mean (or any of the other measures of central tendency) can take - instead of a point estimate we obtain a lower and upper value for the particular measure of central tendency. Engelberg *et al.* (2007) compare point forecasts of annual output growth and inflation from SPF to bounds to see whether the point forecasts and histograms are consistent (see also Clements (2008c, 2008b)).

Clements (2008c) shows that the SPF histograms and probability forecasts can be compared for the Q4 surveys using a bounds approach. For the Q4 surveys, current-quarter probabilities of decline are calculated from the probability distributions, which can be compared to the directly-reported current-quarter forecast probabilities of decline. Given the realized values of output in the seven quarters to the fourth quarter (taken from the appropriate RTDSM), we can infer the year-on-year rate of growth that equates the Q4 level of output with that of the preceding quarter. The implied current-quarter probability of decline from the histogram is then the probability that year-on-year output growth will not exceed this rate. As in the case of calculating means from the histograms, the required calculation could be performed by assuming uniformity within intervals, or by approximating the histograms by normal densities, amongst other methods. But without making any such assumptions about the relationship between the histograms and the respondents' actual beliefs, we proceed as follows.

Consider the following example. Suppose the forecaster attaches probabilities of 0.1, 0.5 and 0.4 to the intervals $[3, 4)$, $[4, 5)$ and $[5, 6)$, with all other bins having zero probabilities. Suppose output will decline if $y < 4.2$. An upper bound on the probability (when all mass in the $[4, 5)$ interval is less than 4.2) is $0.1 + 0.5 = 0.6$, and the lower bound is 0.1, when all mass in the $[4, 5]$ interval is greater than 4.2. Thus the lower bound sums the probabilities of all intervals whose upper bounds are less than the value (here 4.2), and the upper bound includes in this sum the probability attached to the interval containing the value. At the extremes, suppose y lies below the lowest interval. Then the upper and lower bounds on probability coincide at zero. If Y lies above the highest interval, both bounds are 1. Note we are assuming that the upper and lower histogram bins are closed, with the upper limit to the upper bin, and the lower limit to the lower bin, set so that the bin widths are the same as those of interior bins. Bounds calculated in this way satisfy $u, l \in [0, 1]$, and $u - l \in [0, 1]$, where u and l are the upper and lower bounds on the probability.

The first row of table 3 shows that over a third of the probability forecasts lie below the histogram bounds on the probability of a decline, when these directly reported probabilities are taken at face value. Of interest is whether the tendency to round probability forecasts described in section 2 contributes to this finding. The rows labelled M5 to M100 report the percentage of interval forecasts which are inconsistent with the histogram bounds when the intervals are calculated as described in section 2, that is, without exploiting individual response patterns. We find that allowing the grossest form of rounding (M100) reduces the number of interval forecasts that lie below the lower bound on the histogram probability of a decline to around one fifth (20.9%). That

Table 3: Bounds on histogram probabilities of decline and directly-reported probabilities from Q4 surveys (497 forecasts)

Rounding?	% below bounds	% above bounds
As reported	35.2	1.4
M5	32.2	1.2
M10	30.6	1.2
M50	22.5	1.0
M100	20.9	1.0
'Individual'	34.4	1.4

M5 to M100 indicate increasingly grosser degrees of rounding, as described in the text. 'Individual' indicates that the degree of rounding is inferred from individual-level behaviour.

is, for one fifth of the forecasts we find that $p_{j,t_j,U} < l_{j,tj}$ where $p_{j,t_j,U}$ is calculated under M100 and $l_{j,tj}$ is the lower bound on the histogram-derived probability for individual j. It is apparent that allowing for the grossest form of rounding reduces but does not remove the inconsistency documented by Clements (2008c), whereas rounding to a lesser degree (e.g., M5 or M10) has only a relatively small impact.

If we assume that individuals use the same rounding rule for all their forecasts and adopt the scheme described in section 2, then it is apparent from table 3 that replacing the point forecasts by interval forecasts has virtually no effect on the degree of inconsistency. This is inkeeping with the relatively narrow interval forecasts that result under this scheme, as shown in table 2. Note that in calculating the intervals we have used all the current quarter forecasts to infer individuals rounding behaviour, not just the current quarter forecasts made in response to Q4 surveys. This is because the Q4-survey current forecasts resemble the responses to the surveys held in the other quarters of the year in terms of reported forecasts which are multiples of 5 and 10, etc. (not shown to save space). Hence the sample for which we can match current quarter probability forecasts and implied decline probabilities from the histograms appears to be typical all the current quarter probability forecasts, and hence we draw on all those current quarter forecasts.

4 The relationship between the probability of decline forecasts and the point forecasts of quarterly output growth

As well as assessing the impact of rounding on the consistency of the probability of decline forecasts and the probability distributions, we can also investigate the impact of rounding on the relationship between the probability of decline forecasts and the output growth forecasts.

As noted in section 2, respondents typically provide pairs of real output forecasts and probability of decline forecasts for the current quarter, and each of the next four quarters. One would expect these to bear some relationship to each other, although as we explain below, the relationship is likely to depend on a number of factors which are not observable.

Let the probability forecasts be denoted by p_{ith}, which is the forecast by respondent i to the survey dated quarter t, of the probability of a decline in output h-steps ahead. Let x_{ith} denote the forecast of the level of real output. We work with growth rates rather than declines in real output (a forecast of a decline in period $t+h$ would be made if $x_{ith} - x_{it,h-1} < 0$) whereby the percentage change is denoted by $w_{ith,}$, $w_{ith} = 100\left[(x_{ith}/x_{it,h-1}) - 1\right]$. For $h > 0$ we calculate w from the forecasts of the levels.

The quarterly growth rate forecasts would be expected to be negatively associated with forecast probabilities of decline. Further, a respondent with a negative output growth forecast is more likely to report a higher probability of decline the smaller the uncertainty with which that forecast is held. This idea can be formalised by supposing that w_{ith} is the mean of individual i's density forecast of output growth in period $t+h$, and that this density is a member of the scale-location family. Letting W_{t+h} denote quarterly output growth in $t+h$, we have $W_{t+h} \sim D\left(w_{ith}, \sigma^2_{w,ith}\right)$. If we assume that D is the gaussian density, and if $\sigma^2_{w,ith}$ were known, the implied forecast probability of a decline in output ($W_{t+h} < 0$) would be:

$$\widehat{p}_{ith} \equiv \Phi\left(-\frac{w_{ith}}{\sigma_{w,ith}}\right) \tag{1}$$

where $\Phi\left(\right)$ is the cdf of the standard normal. These implied probabilities could then be compared to the reported probabilities. Apart from the normality assumption, the problem with this approach is that in general the individual forecast variances, $\sigma^2_{w,ith}$, will be unknown. One possibility is to fit an ARCH or GARCH[8] model to the consensus forecast errors (as in Bomberger (1996) and Rich and Butler (1998)). The consensus (i.e., average) rather than individual forecast errors are typically used as individual respondents do not file returns to each survey and may file only a small number of responses in total. The drawback is that using the consensus errors supposes that the forecast variance is the same for all respondents.

In order to investigate the relationship between p and w without making such restrictive assumptions, and in order to focus on the potential impact of rounding of the probability forecasts, it seems preferable to estimate the relationship between p and w non-parametrically. Specifically, we use the Nadaraya-Watson kernel smoothing method to estimate the conditional expectation of p given w, $E(p \mid w)$, separately for each forecast horizon h but using all observations on i and t for a given h. We plot the estimated conditional expectation for $h = 0$ and $h = 4$ as the solid lines in figures 1 and 2 respectively (the same estimates appear in each of the four panels of the figures).[9] For the current quarter forecasts there is a clear negative slope, as expected, although the relationship is not linear, with the gradient declining at higher levels of output growth. For $h = 4$ we observe a similar relationship although the probability appears to reach a floor at a higher level, indicating that higher probabilities of a decline in output are associated with forecasted quarterly rates of output growth of around 1 percentage point.

If we do not wish to take the reported probabilities at face value, then we can make the weaker assumption that the reported p is only partially identified (in the sense of Manski (2003)) in that it lies within the interval $[p_L, p_U]$. Using the interval data, Manski and Tamer (2002) show that the identification region for the conditional expectation $E(p \mid w)$, $H[E(p \mid w)]$, is defined by $[E(p_L \mid w),\ \ E(p_U \mid w)]$. We can estimate this region by applying the Nadaraya-Watson kernel method to $E(p_L \mid w)$ and $E(p_L \mid w)$ separately, to give the estimates $\overline{p}_{L|w}$ and $\overline{p}_{U|w}$, say, so that $\hat{H}[E(p \mid w)] = \left[\overline{p}_{L|w}, \overline{p}_{U|w}\right]$. Figures 1 and 2 plot $\overline{p}_{L|w}$ and $\overline{p}_{U|w}$ along with $\overline{p}_{|w}$ (the kernel estimate of $E(p \mid w)$), for four different methods of generating intervals from the reported p values: using individual response patterns, under M10, under M50, and under M100. (M5 is omitted as the results were very similar to using individual response patterns).

Consider figure 1 (figure 2 is essentially the same). It is clear that the estimated identification

[8] See Engle (1982) and Bollerslev (1986).

[9] We use a grid of one hundred equally spaced values of w ranging from the first to ninth decile of the data support of w, and the standard choices of an Epanechnikov kernel function with Silverman's rule of thumb bandwidth.

region $\hat{H}[E(p \mid w)]$ using individual response patterns (top left panel) is sufficiently small that rounding of this type has no effect on the inference we make about the relationship between p and w. If we are not prepared to make the assumption that individuals apply common rounding rules, and instead suppose that the reported p result from one of the grosser forms of rounding (M50 and M100 are shown in the bottom two panels) then the identification region becomes wide as w increases. For example, under M50 there are a sufficient number of relatively uninformative responses of 0,50,100 that $\overline{p}_{U|w}$ roughly bottoms out at w equal to half a percentage point growth. At higher rates of growth we are unable to infer that there is a negative relationship between p and w.

5 Conclusions

We have investigated the potential impact of rounding on the SPF probability forecasts of a decline in real output in two dimensions: in terms of the effect of rounding on the relationship between the probability forecasts and the probabilities of decline imputed from the histogram forecasts; and in terms of the relationship between the probability forecasts and the quarterly output growth forecasts. Because the degree of rounding in the reported probability forecasts is unknown, we have undertaken this investigation under a number of different assumptions about rounding behaviour.

For the SPF probability forecasts we have been able to show that the assumption that individual respondents always engage in the same degree of rounding results in interval forecasts that are sufficiently narrow that in practice rounding can be ignored. If we dispense with this assumption and assume, for example, that probability forecasts reported as point values of 0,50 or 100 reflect 'rounding to the nearest 50', and so replace these point forecasts by the intervals [0,25], [25,75] and [75,100], then the situation changes. Firstly, the percentage of forecasts for which the probability forecasts (represented as intervals) are inconsistent with the imputed probabilities from the histograms is reduced, but not to zero, so that the puzzle identified by Clements (2008c) remains. Secondly, the grosser forms of rounding (such as rounding to the nearest 50) have a greater impact on non-parametric estimates of the relationship between the probability forecasts and output growth forecasts. Using interval forecasts, we are unable to infer that the forecast probability of a decline in output is negatively associated with the forecast quarterly growth rate when the latter exceeds half a percentage point.

Finally, it is worth remarking that the other types of forecasts made by the SPF respondents may be subject to rounding to some degree. While the probabilities attached to the histogram bins may be rounded, in principle the forecasts of the level of real output (from which we calculate the forecasts of the quarterly growth rates) may also have been rounded. It seems likely that rounding is less of a problem for the point forecasts, and moreover the impact of rounding on the histograms should be ameliorated by the constraint that the sum of the probabilities attached to the bins is one.

References

Bollerslev, T. (1986). Generalised autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **51**, 307–327.

Bomberger, W. A. (1996). Disagreement as a measure of uncertainty. *Journal of Money, Credit and Banking*, **28**, 381–392.

Clements, M. P. (2006). Evaluating the Survey of Professional Forecasters probability distributions of expected inflation based on derived event probability forecasts. *Empirical Economics*, **31, No. 1**, 49–64.

Clements, M. P. (2008a). Consensus and uncertainty: Using forecast probabilities of output declines. *International Journal of Forecasting*, **24**, 76–86.

Clements, M. P. (2008b). Explanations of the inconsistencies in survey respondents forecasts. Working paper, Department of Economics, University of Warwick.

Clements, M. P. (2008c). Internal consistency of survey respondents' forecasts: Evidence based on the Survey of Professional Forecasters. In Castle, J. L., and Shephard, N. (eds.), *The Methodology and Practice of Econometrics*: Oxford University Press. Forthcoming.

Croushore, D., and Stark, T. (2001). A real-time data set for macroeconomists. *Journal of Econometrics*, **105**, 111–130.

Davies, A., and Lahiri, K. (1999). Re-examining the rational expectations hypothesis using panel data on multi-period forecasts. In Hsiao, C., Lahiri, K., Lee, L. F., and Pesaran, M. H. (eds.), *Analysis of Panels and Limited Dependent Variable Models*, pp. 226–254: Cambridge University Press.

Diebold, F. X., Tay, A. S., and Wallis, K. F. (1999). Evaluating density forecasts of inflation: The Survey of Professional Forecasters. In Engle, R. F., and White, H. (eds.), *Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive Granger*, pp. 76–90. Oxford: Oxford University Press.

Engelberg, J., Manski, C. F., and Williams, J. (2007). Comparing the point predictions and subjective probability distributions of professional forecasters. *Journal of Business and Economic Statistics*. Forthcoming.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity, with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**, 987–1007.

Faust, J., Rogers, J. H., and Wright, J. H. (2005). News and noise in G-7 GDP announcements. *Journal of Money, Credit and Banking*, **37 (3)**, 403–417.

Giordani, P., and Söderlind, P. (2003). Inflation forecast uncertainty. *European Economic Review*, **74**, 1037–1060.

Keane, M. P., and Runkle, D. E. (1990). Testing the rationality of price forecasts: new evidence from panel data. *American Economic Review*, **80**, 714–735.

Manski, C. F. (2003). *Partial Identification of Probability Distributions*. New York: Springer-verlag.

Manski, C. F., and Molinari, F. (2008). Rounding probabilistic expectations in surveys. Mimeo, Northwestern University.

Manski, C. F., and Tamer, E. (2002). Inference on regressions with interval data on a regressor or outcome. *Econometrica*, **70**, 519–566.

Rich, R. W., and Butler, J. S. (1998). Disagreement as a measure of uncertainty: A comment on Bomberger. *Journal of Money, Credit and Banking*, **30**, 411–419.

Zarnowitz, V. (1985). Rational expectations and macroeconomic forecasts. *Journal of Business and Economic Statistics*, **3**, 293–311.

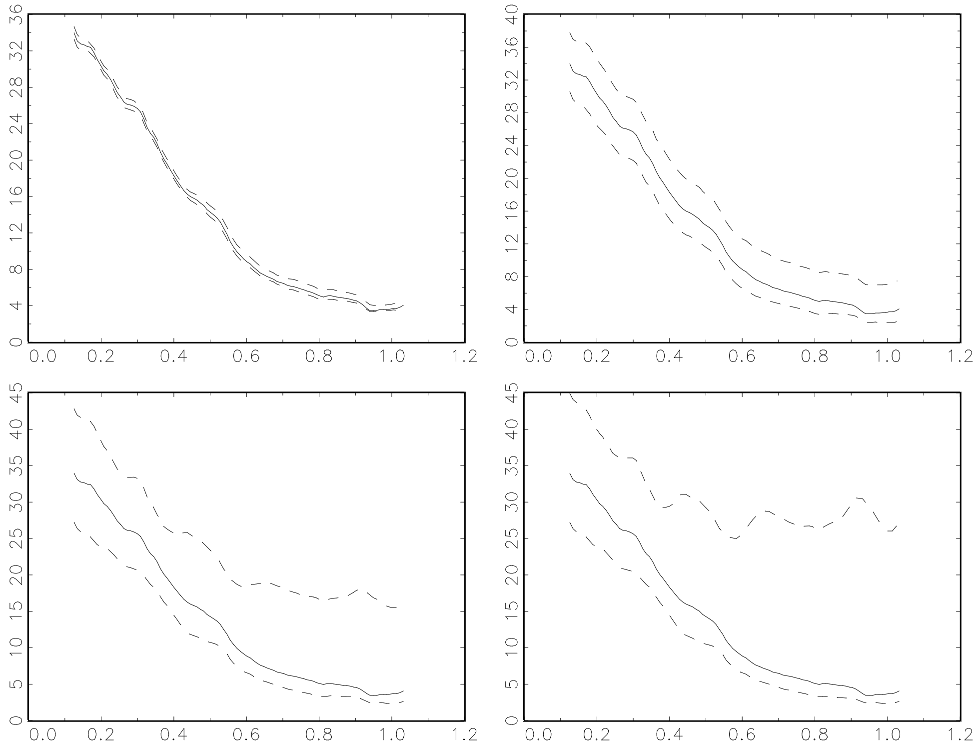

Figure 1: Nadaraya-Watson estimates of the expected value of p given w for current quarter forecasts. The vertical axes are the probabilities of a decline, and the horizontal axes the percentage growth rates. The top left plots estimates for p_L, p and p_U, where $[p_L, p_U]$ are calculated based on individual response patterns; top right is the same but with $[p_L, p_U]$ calculated under M10; for the bottom left and right panels $[p_L, p_U]$ is calculated under M50 and M100 respectively.

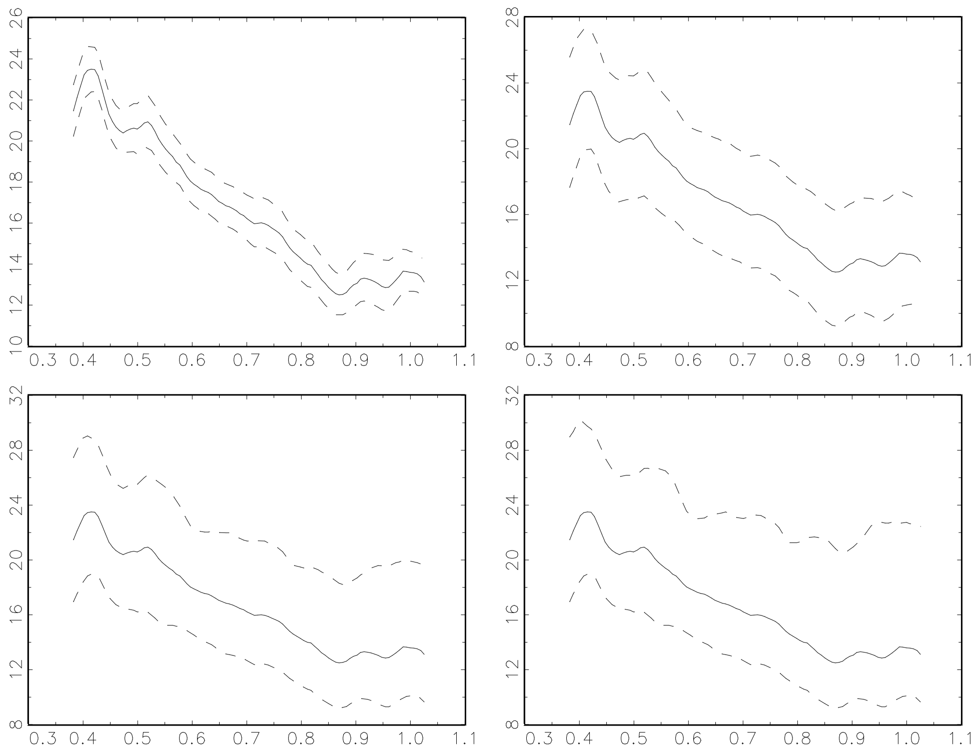

Figure 2: Nadaraya-Watson estimates of the expected value of p given w for four quarter ahead forecasts. The vertical axes are the probabilities of a decline, and the horizontal axes the percentage growth rates. The top left plots estimates for p_L, p and p_U, where $[p_L, p_U]$ are calculated based on individual response patterns; top right is the same but with $[p_L, p_U]$ calculated under M10; for the bottom left and right panels $[p_L, p_U]$ is calculated under M50 and M100 respectively.